Shortcut Shot Detection Based on Compressed Video Bitstream

Kwang Ryol Ryu and Young Bin Kim, *Member, KIMICS*

Abstract—The shortcut shot detection based on MPEG compressed video bitstream is presented in this paper. The detection algorithm is used the video picture frame from MPEG compressed video directly not to be decompressed the original image. For shortcut detection, I and P frame of MPEG video bitstream are classified. The changing scene cuts at I pictures are detected by the decompressed DC image and scene cuts at P picture frame by monitoring the percentage of intra-macroblocks per P picture are detected. Experimental results using test video bitstream QVGA results in average 92% detection rate, searching time is taken around 4.5 times faster in comparison with changing scene shot detection algorithm which is decompressed the compressed bitstream.

Index Terms—Shortcut shot detection, changing scene, video picture type, macroblock, MPEG, bitstream

I. INTRODUCTION

A demand of sequential indexing and fast searching in mass multimedia information is increased progressively. In order to fulfill the request these video streaming method needs to partitions for scene by scene, the shot detection technique on a changing scene is required that search for the shot image position in the sequence of video frames wholly. There are two type different techniques for a changing scene, frame base or compressed video base. This method of frame base enables to detect a changing scene clearly in comparison with the detection method based on the compressed video data because of using the special value on the frame like histogram. However it takes a much time to calculate the operations for detecting of the changing scene. An objects and camera moving fast, and variation of illumination instantaneously, and the gradual changing scene is not sensitive. Miller and Mai proposed the method using edge for overall a changing scene detection in cut, fade-in, fade-out, dissolve, wipe. Zabih et al proposed an edge-based method, which detects the appearance of intensity edges that are distant from edges in the previous frame and appears to be more accurate at detecting and classifying a changing scene detection points that are difficult to detect with intensity histograms. Hampapur and Shaharay et al has proposed model base method for problem of histogram level. [1~6]

There are several other advantages using the base on the compressed video bitstream method, first, there are less number of data in compressed domain verse pixilated domain which reduce the overhead running an algorithm. Secondly, the small block of data files are processing in compressed method will reduce total processing time. Lastly, these data doesn't need to recover on the original format and also reduce total processing time. However, this format doesn't use a touch original frame futures, picture quality is not as good as the other method. Arman et al proposed segmenting JPEG video with a difference metric based on the correlation between DCT (Discrete Cosine Transform) coefficients of consecutive frames. Zhang et al proposed an approach to exploit the information of motion vectors between consecutive P (Predicted picture) frames to detect the camera motion. Nakajima has developed a fast changing scene detection approach by calculating the correlation of DCT coefficients of I (Intra-picture) frames. However changing scene in P or B (Bidirectional predicted picture) picture frame did not detect. Also, Meng has proposed for abrupt shot boundary detection is used the moving vector distribution, and the gradual shot boundary detection is used the DC coefficient difference of DCT. [7~11]

This paper is proposed a shortcut shot detection method using I and P picture frame in MPEG compressed video bit stream for changing scene shot detection. The expands are that I and P frames are classified picture type to P and I frame, scene cut at I pictures are detected by decompressed DC image and changing scene shot at P picture frame by monitoring the percentage of intra-macroblocks per P picture is detected with experiment. That makes the shot detection to leads cut position.

II. SHORTCUT SHOT DETECTION

A. Shortcut Detection Processing

The picture types on video streaming can be distinguish based on format of compressed data by I, P, B for GOP(Group Of Picture) or I, P. By any chance, if the GOP is without B picture type the Zhang method can not be use to produce the frames. In order to produce scene conversion the DC picture will be decompressed

Manuscript received July 9, 2007.

Kwang Ryol Ryu and Young Bin Kim are with the Department of Electronics Engineering, graduate school of Mokwon University, Daejeon, 302-729, S. Korea (Tel: +82-42-829-7651, Email: ryol@mokwon.ac.kr)

by DCT coefficients base on before and after I picture types. However, scene conversion which makes P type can to be produced using I picture type by itself. When threshold value in macroblock is higher the frame conversion in P picture type use macroblock counts to produce using shot mode. The whole processing of shortcut shot detection is shown in Fig. 1 by based on moving compensation of MPEG. The compressed video bitstream MPEG is inputting is divided into I and P picture types. I picture is analyzed DC image from DCT coefficient and is operated histogram, and it is compared with the preceding I picture histogram. The divided P picture type is calculated the number of macroblocks and is applied to threshold value.

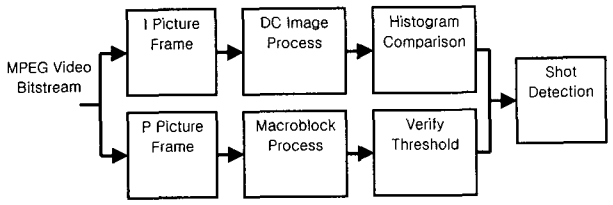


Fig. 1 Shortcut shot detection procedure

B. Moving compensation

The MPEG video streaming is classified audio and video signal on the frame header. The video signals are organized by number of individual picture frame. The Picture means, non-motion video, and GOP are the motion picture. The MPEG formed with four different picture frames, these are distinguished by header information. I frame does not compensate the movement but encode the corresponded frame by DCT. P frame operates DCT to the difference in the moving compensation of I frame or the P frame. B frame compensates the movement like the P frame. But, unlike the P frame, it compensates from the reference frames (I or P). GOP configuration is I frame and P frame, or it can include I frame, P frame, and B frame. In the GOP configuration, at least one I frame is required. Fig 2 is shown the video sequence having the N=15 and M=1. In the figure, one GOP includes 14 P frames. Frames of P5~P12 are expressed as dot lines. I frame of gray becomes the start frame of each GOP. P1 becomes the second frame which is revived by referring the previous I frame.

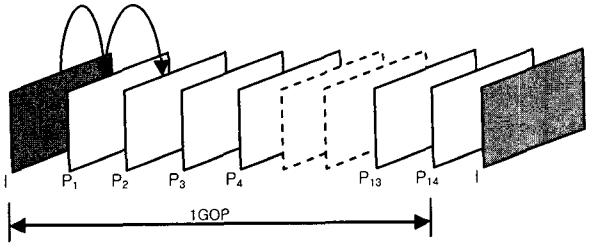


Fig. 2 I and P frames of GOP configuration

C. I frame Shot detection

I frame can be reconstituted without reference because it is encoded not removing the redundant information with the other frames. DC images are decompressed by using DC term of DCT coefficient on the compressed video bitstream. DC image is small by 1/64 of the original image. But, it has the important characteristics of the original image. These characteristics can be used in the image analysis effectively. In case of using DC image, it can process the image quickly as an advantage. Also, it is easy to decompress the DC image in I frame, but it is difficult to decompress the DC image in P, B frames and can be sensitive because of its small size as a disadvantage. To decompressed DC image intensity of light is applied to filter to simplify itself. Even the filtering process makes the simplified image to change the intensity abruptly in the original image. By such way, images are decompressed in I picture frame is detected the changing scene shot using the difference of the histogram.

D. P frame Shot detection

In case of P frame, the forward moving vector is detected to be based on the previous P picture frame and P frame is reconstructed by the detected moving vector. The changing scene shot is extracted by the variation of macroblock in neighbor P frame. A macroblock comes from GOP. A group of picture is a set of picture in contiguous frames sequence. Each picture is divided into slice. Slices can be as big as the whole picture and as small as a single macroblock. A slice header contains information for its position within a picture and a quantizer scale factor, between one and 31, that can be used by the decoder to dequantize the coded DCT coefficients. A slice has macroblock with 16x16 pixels. Macroblock consists of 4 blocks with hue and 2 blocks with luminance. Macroblock is reconstructed using MPEG header information. The components of 8x8 macroblock consist of luminance(Y), red color: Cr and blue: Cb. The decoding sequences priority is ordered by 4 luminance elements first, red and blue element secondly.

The MPEG video bitstream is encoded by macroblock unit. A macroblock is encoded properly owing to depend on each frame variation. That is, if the moving compensation is ineffective caused by changing scene or intra mode encoding in the corresponding frame is more effective than the moving compensation result, then macroblock is encoded in intra mode regardless of the kinds of the frame. Frame which has this number of macroblock is plentiful is detected in a candidate shot because that frame is possible to cause a changing scene highly. The detected candidate shot θ is involved a key to choose the threshold of historgram difference. Equation (1) is the condition to detect the candidate shot.

$$TH_{\max} : MB_{\max} \le \theta$$
$$TH_{\min} : MB_{\min} \le \theta \le MB_{\max} \quad (1)$$

The macroblock $MB_{\min}$ is the minimum number of macroblock encoded in intra mode, has 40% of the total candidate macroblocks by experiment. The macroblock

MB_{max} is the maximum number of the macroblock, has 70% of the total candidate macroblocks. Threshold value TH_{max} and TH_{min} to detect changing scene shot depends on the number of macroblock. Threshold value TH_{max} is that the candidate macroblocks θ is higher than the number of macroblock of P frame MB_{max}. That given low threshold value in order to detect in case of occurring the previous changing scene. TH_{min} is that if the candidate macroblokks θ is great than and equal to MB_{min}, and less than and equal to MB_{max}, is given high value to detect the changing scene having the similarity relevant to the previous scene. Fig. 3 is shown the process of detecting the number of macroblock in P_2 and P_3, and compare with both for detecting the candidate changing scene.

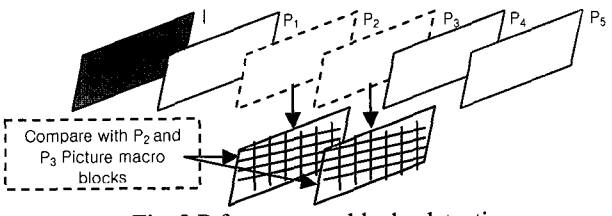


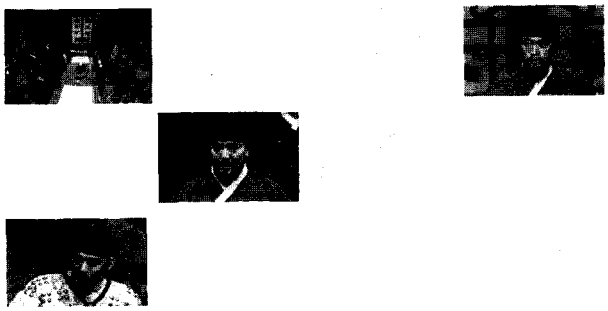
Fig. 3 P frame macroblocks detection.

III. EXPERIMENTAL RESULTS

In order to evaluate the changing scene shot detection proposed method, three MPEG-4 video bitstreams are used. Test bitstream has a compressed GOP structure in N=15 and M=1. The test video image is applied to QVGA (Quarter Video Graphic Array) 320x240 pixels in size such as DMB (Digital Multimedia Broadcasting). The algorithm is processing in two distinguished steps. First, the method detects the picture type in frame. Second, Extract the standard data for a shot detection and verify whether occur a changing scene or not. DC image is decompressed in I frame, the number of DCT macroblocks is detected in P frame. I frame is detected each one every GOP. Hence, The DC image is decompressed by detecting I frame in the first GOP and the second GOP respectively. The decompressed each DC image should be normalized by the histogram not to be sensitive to the change of the light. The histogram difference is calculated in the DC image. If the difference is 0.6 over than is detected as a changing scene shot.

P picture sequence is enumerated between I pictures as shown Fig.2. The number of macroblocks for all P ictures should be calculated. If the difference between the number of the previous P frame macroblock and the number of the present P frame macroblock is more than the threshold value, then is detected in the changing scene. Fig. 4 is shown the result that is progress for shot and cut detection in the sequence of video frames.

a. Sequence of video frames



b. Detected shot and cut frames

Fig. 4 Frame sequences and detected shot frames and positions

Table 1 is shown the result of scene change detection using DC image to I frame. As it is shown, when I picture is used to detect scene change in Video 1, Video 2, Video 3, the rate of detecting are 98%, 92% and 85%. The accuracy of detecting scene change in the Video 3 is low. The reason is that Video 3 is fast and large movement relative to the Video 1, and Video 2.

Table 1 Shot detection at I picture

Test Video	Correct Detection	False Detection	Total Frames
Video 1	75	2	9000
Video 2	79	7	9000
Video 3	102	18	9000

Table 2 is shown the result of the detecting using the macroblock to P frame. In the test video, the results of the scene changing shot detection are 98%, 96%, 93%.

To detect of scene changing shot, the threshold values are 40% and 70%. In case of high threshold value, the precision increases.

Table 2 Shot detection at P picture

Test Video	Correct Detection	False Detection	Total Frames
Video 1	156	3	9000
Video 2	180	6	9000
Video 3	264	17	9000

Table 3 is shown the result of comparing the detecting

speed depending on the method. The detecting ability of the scene changing shot for performing speed is measured by input of Video 1 in bit stream. That is, restore the compressed bit stream data to the original image and detect the scene changing shot by using the histogram difference of reconstructed image. Method 2, 3 are the basic scene change detecting methods. Method 2 is using the DC coefficient of I frame. Method 3 is the proposed way. As the result of the experiment, Method 2 and 3 using the compressed data are around 4.5 times faster than the Method 1.

Table 3 Speed comparison of shot detection

Test Video	Method 1	Method 2	Method 3
Video 1	3.2(f/s)	15.6(f/s)	13.4(f/s)

IV. CONCLUSIONS

The shortcut shot detection method to be used I and P picture type on video bitstream of the compressed MPEG is proposed in this paper. The method doesn't need decode the compressed MPEG bitstream into the original image for the detection. On I picture, the shot is detected by comparing with the histograms between DC image is restored using DCT coefficient and the restored adjacent DC image for GOP. On P picture referred I picture, the shot is extracted by the threshold value calculating the number of macroblock of P picture. The minimum threshold has about 40% of the total macroblocks and the maximum 70% of the candidate shot of macroblocks by experiment. That results lead detection rate to be 92% on I picture and 96% on P picture, and also, 92 % on the test video bitstream QVGA. The proposed method enables to be faster as around 4.5 times in comparison with the conventional method and algorithm which the shot detection is used by decompressing the compressed MPEG frame. In case of failing to detect the shot at the same macroblock on the compressed picture, the method is required to integrate and optimize the algorithm in future.

REFERENCES

[1] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Prentice-Hall, 2002.

[2] Ricahrd S. Shernock Regis J. Crinon, Michael A. Dolan, John R. Mick Jr., Data Broadcasting, McGraw-Hill, 2001.

[3] Joung-Dae Jeon, Seong-Il Cheong, Sang-Ryul Ryu, Sung-Ho Kim, "Cut Detection Using Macroblock and Frame Information", The Korea Information Science Society, vol.28. No.2, pp.271-273, 2001.10.

[4] In-Guk Jeong, Hwa-Sun Lee, Jong-Ho Nang, Gyung-Su Kim, Myung-Hwan Ha, Byung-Hui Jeong, "A New Shot Change Detection Scheme Using Color Histogram and Macroblock Information of MPEG Video Stream", The Korea Information Science Society, vol28. No 01, pp.418-420, 2001.4.

[5] Hang-Bong Kang, "Content-based Shot Boundary Detection from MPEG Data using Region Flow and Color Information", The Journal B of The Korea Information Science Society, vol.27 No.4, pp.402-411, 2000.4.

[6] Seong-Whan Lee, Young-Min Kim, Sung-Woo Choi, "Fast Scene Change Detection using Direct Feature Extraction from MPEG Compressed Videos", IEEE TRANSACTIONS ON MULTIMEDIA, Vol.2, NO.4, pp.240-254, 2000.12.

[7] Ali M Dawood, Mohammad Ghanbari, "Scene Cut Detection form MPEG Video Stream Coded without B Pictures", IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING, Vol.3 No.1, pp.1645-1648, 2001.

[8] F. Arman, A. Hsu and M.-Y. Chiu, "Feature Management for Large Video Databases", in Storage and Retrieval for Image and Video Databases, Vol. SPIE 1908, pp.2-12, 1993.

[9] F. Arman, A. Hsu and M.-Y. Chiu, "Image processing on Compressed Data for Large Video Databases", Proc. ACM Multimedia'93, Jun., pp.267-272, 1993.

[10] H. J. Zhang, C.Y. Low and S. W. Smolia, "Video Parsing and browsing using compressed data", Multimedia Tools and Applications, Vol.1, No.1, Mar., pp.89-111, 1995.

[11] Young-Bin Kim, Kwang-Ryul Ryu, Robert J. Sclabassi, "Changing Scene Detection using Histogram and Header Information of H.264 Video Stream", Proceedings of the Korean Institute of Maritime Information and Communication Sciences Conference, pp.197-200, 2006. 06.

Kwang Ryol Ryu
Ph. D. is currently a professor at the Division of Information & Electronics & Imaging Engineering at Mokwon University, works as a visiting Professor at the Dept. of Neurological Surgery at University of Pittsburgh Medical Center in USA and now vice president of KIMICS. His areas of research include DSP, Imaging and Vision, and Bio-Medical Electronics.

Young Bin Kim
Has been Ph. D. student at Mokwon University. He received his undergraduate education and M.S. in Electronics Engineering from the Mokwon University in Korea. His areas of research include Digital Image Pro-cessing and Video Application